Data Dictionary ("Data Model")
7/16/03, Roberts and Rezapkin suggestions

This is pretty close to what we are using now in our implementation. We have included the SPASE categories where they are well-defined. We have not otherwise tried to be completely faithful to the current SPASE definitions given that much is subject to change. The dictionary presented here includes very nearly all the terms in the CDPP Inteoperability table of terms used at data centers; the major exception is that we have not spelled out "Location" in the sense that is required for stating which precise region a spacecraft is in at a given time.

Our current charge was to take the next cut at "Parameter/Entity" and that is listed below. We have changed from "Parameter" to "Physical Quantity" since the convention in much of physics is that a "parameter" is a constant in an equation, as opposed to the variables in it. We believe the new term better reflects the intended idea and is in accordance with some earlier data dictionaries. The term "Property" has also been used (e.g., by Chris Harvey), and this would work too.

Note that "Data Product," "Header," and "Time Series," are markers that do not have values but simply serve to organize the list. In principal, each element (**BOLD**) can be seen as independent; multiple elements and their values determine the characteristics of a product or quantity within the product. An implementation of the dictionary could specify "business rules" that would restrict the use of particular elements according to some level of ontology; this can get messy and hard to maintain, so we are going to see how much we can avoid it. It is also possible to limit the likelihood of ill-formed queries by restrictions built into the user interface. The listing below is given hierarchically for clarity.

The Header information locates the product in the data universe, and the Time Series information gives the contents of the product.

The list of Entities is based on the natural taxonomy of the physical world, but also corresponds to entities in known products. We do not have good examples, yet, of products with things like aerosols in them, so we have simply included these for completeness. We have defined "Atom" to include neutral atoms, and protons and alphas are in effect listed twice, once explicitly, and once by NZQ.

The physical quantities are the two classical fields (electric and magnetic) and the associated electric potential, number/energy fluxes to describe distribution functions of particles (differential by bins), the first four moments of a particle distribution function, the (perhaps too vague) intensity, counts, and various derived quantities. It remains to be determined if we want a list of all the derived quantities contained in all relevant products; what appears below covers most (all?) of the many cases we have examined. Wave fields are described either as time series of fields, or using the qualifier "spectral density," although products use terms such as "peak electric field" in pretty vague ways. Our intention in such cases is to determine what was really meant and to give a better description than the original.

Qualifiers of Physical Quantities can be combined, e.g., the rms of the x component. The field-aligned component of the proton heat flux vector would be described as the parallel component of the heat flux of protons in magnetic coordinates.

Black **boldface** specifies the element. Blue text is descriptive, but does not indicate a type of metadata. Green text indicates the form and content of the metadata.

DATA_PRODUCT Represents the content of a data product that may contain many physical quantities for many entities, sometimes from multiple instruments.
- o **HEADER** General product specific information.
  - **PRODUCT_DESCRIPTION** (string) - Text description of a product.
  - **REPOSITORY** (identifier) - Identifies the repository where the product is located.
  - **REGION** (enumeration) – Indicates the region(s) occupied by an observatory in the course of its mission. [Intended to be an orthogonal categorization of observatories, not a "location" for a measurement.]
    - Heliosphere
    - Solar Wind Near Earth
    - Solar Wind/Magnetosphere
    - Magnetosphere
    - Geosynchronous
    - Near Earth
    - Ground-based
  - **PROJECT** (enumeration) – A list that lumps together things like all Cluster or GOES spacecraft, or a magnetometer chain.
  - **OBSERVATORY** (enumeration) - Identifies the name of the spacecraft or ground station from which the data originated, given by conventional names. Sometimes the same as "Project."
  - **MEASURMENT_TYPE** (enumeration) - Identifies the type of content in the product. Roughly corresponds to common instrument types.
    - Magnetic Field
    - Electric Field
    - Radio and Plasma Waves
    - Thermal Plasma
    - Ion Composition
    - Energetic Particles
    - Radio Sounder
    - Imager
  - **INSTRUMENT_NAME** (enumeration) – Identifies the instrument using the standard abbreviation. Many aliases?
  - **TIME_RANGE** (time range) The overall range in which data are available for this product. [Gaps may occur.]
  - **PROCESSING_LEVEL** (enumeration) - Identifies the processing level of the product.
    - Raw
    - Uncalibrated
    - Calibrated
    - Derived
  - **USE_AND_ACCESS** (string) – Identifies, e.g., browse or proprietary data.
  - **FORMAT** (enumeration) Roughly, representative form. [Note that image formats are for Image Products: a separate top level category].
    - CDF
    - HDF
    - ASCII

NetCDF
IDFS
UDF
FITS
Custom binary

- o **TIME_SERIES**. A set of numbers contained in the product, each associated with a timestamp and a uniform unit, or a set of numbers that are the ratios of such quantities.
    - **PHYSICAL_QUANTITY** (enumeration) - Identifies the measured property ("parameter"). Roughly, the label of a column or columns in a table of values (may need one or more Qualifiers to specify a column).
        Magnetic Field
        Electric Field
        Electric Potential
        Number Density
        Mass Density
        Velocity
        Temperature
        Thermal Speed
        Heat Flux
        Number Flux [differential specified by using bins]
        Energy Flux
        Intensity
        Counts
        Time of observation
        Position (of observatory)
        Observed position (where the quantity has the value)
        (Below are "derived")
        Index [explicit list? AE, AL, AU, Kp, Dst, PCI, Ap, Cp, C9, ap…]
        Pressure
        Plasma Beta (ratio of thermal to magnetic energy)
        Sonic Mach Number
        Alfvénic Mach Number
        Cross Spectrum [this is many things; includes Stokes parameters]
            FREQ_RANGE (numeric range) – Frequency range of a quantity.
                - **UNIT** (unit) - Identifies the unit of the range.
            POLAR_ANGLE_RANGE (numeric range) – Polar angular range of a quantity.
                - **UNIT** (unit) - Identifies a unit for the range.
            AZIMUTHAL_ANGLE_RANGE (numeric range) – Azimuthal angular range of a quantity.
                - **UNIT** (unit) - Identifies a unit for the range.
            ENERGY_RANGE (numeric range) – Energy range of a quantity.
                - **UNIT** (unit) - Identifies the unit for the range.
            NUM_FREQUENCY_BIN (number) – Number of frequency bins.
            NUM_ANGLE_BIN (number) – Number of angle bins.
            NUM_ENERGY_BIN (number) – Number of energy bins.
    - **QUALIFIER** (enumeration) - Characterizes a quantity more completely. May be used together for greater specificity.
        Vector
        Component
        Magnitude

Spectral Density
Perpendicular
Parallel
Variance
Peak
Average
Deviation
Root Mean Square (Sigma)
Variance
Uncertainty
Polarization
X
Y
Z
R
Theta
Latitude
Phi
Longitude
T
N

- **COORDINATE_SYSTEM** (enumeration) - Identifies the coordinate system.
    GSE
    GSM
    GEO
    GEI
    HGI
    HSE
    RTN
    Carrington (rotating HG)
    BL
    Spacecraft
    Magnetic
- **UNIT** (unit) - Identifies a unit of measurement, one for each quantity.
- **TIME_SERIES_DESCRIPTION** (string) – Text description of a time series.
- **KEYWORD_IDENTIFIER** (string) – Word used in the file (e.g., in HDF) to identify the time series.
- **ENTITY** (enumeration) – Identifies the physical particle or basic entity whose properties are being measured. Represents the physical entity whose properties are being measured.
    Field
    Electron
    Proton
    Alpha
    Atom
    Molecule
    Dust
    Aerosol
    Cluster
    Photon
    Spacecraft
        NZQ Represents the atomic particle specification.

- o **ELEMENT_NUMBER** (number) – Specifies element number.
        - o **CHARGE_NUMBER** (number) – Specifies charge number.
        - o **ISOTOPE_NUMBER** (number) – Specifies isotope number.
- **RATIO** Represents the ratio of two entities.
    - **NUMERATOR** Represents the numerator entity in a ratio.
        - o **ENTITY** (<u>See entity element above, hierarchy repeats</u>)
        - o **PHYSICAL_QUANTITY** (<u>See above</u>)
    - **DENOMINATOR** Represents the denominator entity in a ratio.
        - o **ENTITY** (<u>See entity element above, hierarchy repeats</u>)
        - o **PHYSICAL_QUANTITY** (<u>See above</u>)
- **RESOLUTION** (number) - Specifies the number of seconds between timestamps within a time series.